Statistical analysis plan

Statistical analysis plan (SAP) for clinical outcomes in:

Stratified care physiotherapy integrated with eHealth versus usual primary care physiotherapy in patients with neck and/or shoulder complaints

Publication date:	05-12-2022
Document version:	1.0
Trial registration:	Netherlands Trial Register: NL8249 & ClinicalTrials.gov: X

This document is a supplement to the trial protocol[1] and comprises a statistical analysis plan for the articles: 'Effectiveness of stratified care integrated with eHealth versus usual primary care physiotherapy in patients with neck and/or shoulder complaints: a cluster randomized controlled trial' and 'Cost-effectiveness of stratified care integrated with eHealth versus usual primary care physiotherapy in patients with neck and/or shoulder complaints: a cluster randomized controlled trial'. CONSORT statement and the extension for Cluster Trials will be used to report the findings.[2] This statistical analysis plan is prepared in accordance with guidelines for statistical analysis plans in clinical trials.[3]

Table of contents

Introduction

Background and rationale

Neck and shoulder complaints are common in primary care physiotherapy. These patients experience pain and disability, resulting in high societal costs due to, for example, healthcare use and work absence. Content and intensity of physiotherapy care can be matched to a patient's risk of persistent disabling pain. Mode of care delivery can be matched to the patient's suitability for blended care (integrating eHealth with physiotherapy sessions). It is hypothesized that combining these two approaches to stratified care (referred to from this point as Stratified Blended Approach) will improve the effectiveness and cost-effectiveness of physiotherapy for patients with neck and/or shoulder complaints compared to usual physiotherapy.

Objectives

Our primary research aim is to investigate the clinical effectiveness of the Stratified Blended Approach for patients with neck and/or shoulder complaints on a combined measure of pain and disability over 9 months, compared to usual physiotherapy care.

Our secondary aims are twofold:

- to investigate the effectiveness of the Stratified Blended Approach for patients with neck and/or shoulder complaints on pain intensity, health-related quality of life, illness perceptions, physical activity, self-management skills and self-perceived effect over 3 and 9 months; and exercise adherence and satisfaction at 3 months, compared to usual physiotherapy care;

- to investigate the cost-effectiveness and cost-utility of the Stratified Blended Approach for patients with neck and/or shoulder complaints over 9 months, compared to usual physiotherapy care.

Trial methods

Design and randomization

A multicenter, pragmatic, two-arm, parallel-group, cluster randomized controlled trial (cRCT) will be conducted. Physiotherapists will be recruited from primary care physiotherapy practices across all regions of the Netherlands. After recruitment, participating primary care physiotherapy practices will be randomized to either offer the Stratified Blended Approach or usual physiotherapy care by a computer-generated random sequence table generated using SPSS, using 1:1 allocation. The number of physiotherapists per participating practice (dichotomized into '1' and '>1') will be used as a stratification variable. Although individual physiotherapists are identified as clusters, physiotherapy practices will be the unit of randomization in order to prevent contamination between physiotherapists within one practice.

Sample size

We aim to reach a sample size of 238 patients. The full sample size calculation is provided in the trial protocol.[1]

Framework

All outcomes are analyzed for superiority of the Stratified Blended Approach arm compared to the usual physiotherapy arm.

Statistical interim analysis and stopping guidance

Interim analyses are not planned.

Timing of final analysis

Analyses will take place after locking the database. The database will be locked after the last patient completes the final follow-up questionnaire and after data are checked on outliers and missing data (anticipated December 2023).

Timing of outcome assessments

The first assessment will consist of a digital questionnaire and an accelerometer that will be sent to the patients by mail and patients will be asked to wear the accelerometer for five consecutive days. Patients will be asked to complete the first questionnaire within one week after starting physiotherapy treatment. Outcomes will be measured again at 3, 6, and 9 months after the first digital questionnaire was completed. Schedule of enrolment, interventions and assessment is provided in our trial protocol (Table 2).[1]

Statistical Principals

Confidence intervals and *P* values

All analyses are described in this statistical analysis plan and are therefore considered *a priori* analyses. If applicable, *post hoc* analyses will be identified as such in the articles.

For all analyses, a two-tailed significance level of $p< 0.05$ is considered to be statistically significant. All confidence intervals presented will be 95% and two-sided. The assumption of normality will be checked with Q-Q-plots and histograms. The interquartile range will be reported for skewed data.

Statistical analysis will be performed using IBM SPSS 27 or statistical package STATA. During the analyses, the researchers will be blinded to group allocation until the entire analysis will be completed.

Adherence/compliance and Protocol deviations

Compliance to the smartphone app with e-Exercise modules in the Stratified Blended Approach arm will be assessed by quantitative data on the usage. These data will automatically be stored on the backend of the app. Additionally, all patients will be asked in the first follow-up questionnaire whether they received and used an app or paper-based workbook as part of their physiotherapy treatment. Participants will be considered compliant to the e-Exercise modules (app) if they log in once a week in 67% (low risk: 2 weeks over 3 weeks) or 75% (medium/high risk: 9 weeks over 12 weeks) of the total amount of weeks. Participants will be considered to comply to the paper-based workbook if they self-report that they used the workbook at T1 (3 months). Descriptive statistics on the percentage adherent (plus N, mean, SD, median, minimum, maximum) will be summarized per arm.

Because of the pragmatic design of the trial, there are no protocol deviations that might impact the analyses. An overview of the content of physiotherapy sessions & number of physiotherapy sessions will be described for both arms.

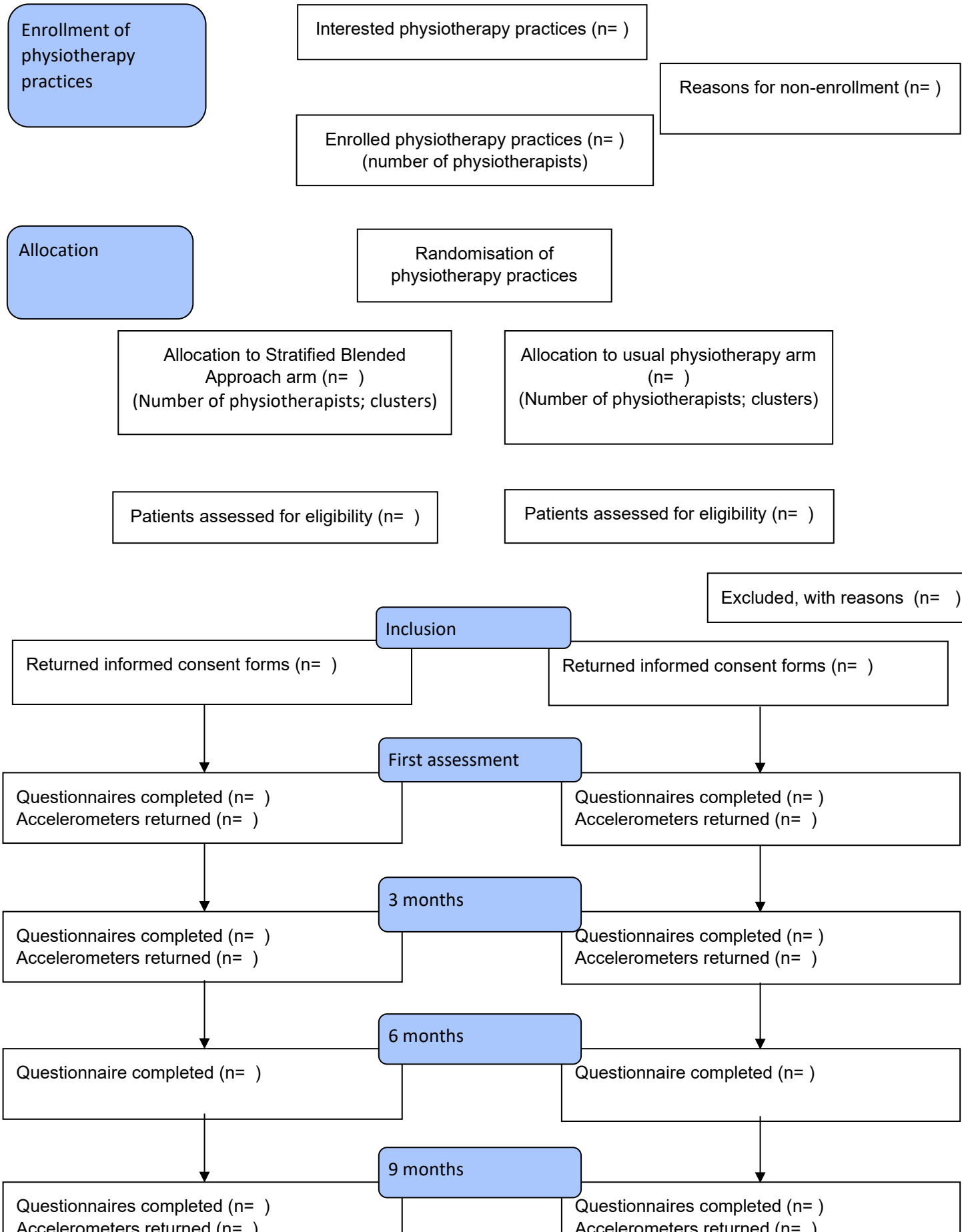
Analyses populations

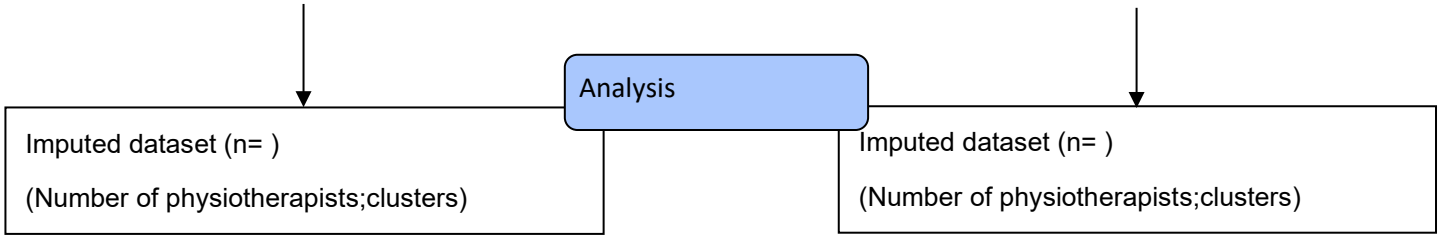
All main analyses will be performed according to the 'intention-to-treat' principle. Missing value analyses or imputation will not be performed, because primary analyses will be linear mixed models (LMM).

Additionally, per protocol analyses will be carried out with people that complied to the paper-based workbook or the e-Exercise modules, compared to the control arm.

Trial Population

Screening data, eligibility, recruitment and loss to follow-up





'Baseline' patient characteristics

Patients will be described with respect to age, sex, education level, type of complaints, duration of complaints, BMI, risk of persistent disabling pain, smartphone or tablet access, amount of apps used weekly, co-morbidities, physical activity level (MVPA, average hours per day), pain and disability score, pain intensity, health-related quality of life, illness perceptions and self-management skills. T0 data of the intervention arm will be shown in a table for all cases with baseline measurements, for both treatment arms. T0 data will also be shown for the 6 subgroups in a supplementary table. Data will be reported as number (percentage of participants), as median (IQR) or as mean (SD).

Table 1. Patient characteristics at T0

Variable	**T0 value**	
	Stratified Blended Physiotherapy	**Usual physiotherapy**
No. of respondents		
Age, years		
Sex, female (%)		
Education level		
Low		
Middle		
High		
Type of complaints (%)		
Predominantly Neck		
Predominantly Shoulder		
Duration of physical symptoms		
<3 weeks		
3 – 12 weeks		
>12 weeks		
BMI, kg/m2		
Risk of persistent disabling pain (using the STarT MSK Tool)* (%)		
Low		
Medium		
High		
Smartphone or tablet access, yes (%)		
Smartphone or tablet usage		
None		
1 – 3 apps per week		
4 – 10 apps per week		
>10 apps per week		

No. of comorbidities	
0	
1	
≥2	
Combined region-specific pain and disability score[†]	
Health-related quality of life[‡]	
MVPA (average minutes per day)[§]	
Pain intensity[¶]	
Illness perceptions[#]	
Self-management skills[**]	

Analysis

Outcome definitions

Effectiveness

Primary outcome

- The primary outcome is the combined region-specific pain and disability score over 9 months follow-up, assessed by the Neck Pain and Disability Scale (NPAD)[4–8] for patients with primarily neck complaints and by the Shoulder Pain and Disability Index (SPADI)[9–13] for patients with primarily shoulder complaints. A higher total score (0-100 for both outcome measures) indicates increased pain and functional limitations.[4–13]

Secondary outcomes

- The average neck and/or shoulder pain intensity in the last week will be measured with an 11-point Numeric Rating Scale (NRS) (0 = no pain; 10 = worst pain imaginable).[14,15]
- Health-related quality of life will be measured with the 36-Item Short Form Health Survey (SF-36). The questionnaire consists of eight subscales (physical functioning, role limitations due to physical health, role limitations due to emotional problems, energy/fatigue, emotional well-being, social functioning, pain (over the last 4 weeks) and general health). Scores for each subscale will be calculated (0-100). Higher scores indicate a better health-related quality of life.[16–20]
- Illness perceptions will be measured with the Brief Illness Perception Questionnaire (IPQ-K).[21–23] This questionnaire is an eight-item scale designed to assess cognitive and emotional representations of illness on an ordinal scale (0-10).[21–23]
- Patients' self-management skills are assessed with the Dutch version of the short form Patient Activation Measure (PAM13-Dutch).[24,25] The PAM 13-Dutch is a reliable 13-item instrument and assesses patient (or consumer) self-reported knowledge, skills and confidence for self-management of one's health or chronic condition. The answering categories per item are 4-point Likert scales, ranging from totally disagree to totally agree

and 'non applicable'. A higher score (range 1-100) indicates a higher level of self-management.[24,25]

- Physical activity will be measured with an Actigraph accelerometer.[26,27] The Actigraph accelerometer is a reliable tool for measuring physical activity in adults. Participants will be instructed to wear the accelerometer on their waist for five consecutive days, except when sleeping, showering, bathing or swimming.[26,27] Average amount of moderate or vigorous physical activity (MVPA) per day will be calculated. Data will be eligible if patients had worn the meter at least 3 days, for 8 hours or more.[28] PA thresholds of Freedson et al.[29] were used to distinguish sedentary activity and light, moderate, and vigorous PA. Moderate activity and vigorous activity will be summed and divided by the number of wearing days to calculate a PA score in minutes per day.
- Exercise adherence will be measured with the Exercise Adherence Rating Scale (EARS).[30] The EARS is a 6 item self-reported questionnaire with items scored on a 5-point Likert scale (0 = completely agree; 4= completely disagree). A higher score (0-24) indicates better adherence to prescribed home-exercises.[30] This outcome will not be measured at 9 months follow-up.
- Global perceived effect will be measured with the 7-point Likert global perceived effect score (GPE).[31,32] Categories 1 (very much improved) to 3 (a little improved) are classified as 'improved'. Categories 4 (no change) to 7 (very much worse) are classified as 'not improved'.[31,32]
- Satisfaction with treatment outcome will be measured with an 8-point Likert scale question: 'All things considered, how satisfied are you with the results of the treatment for your neck and/or shoulder complaints? (1 = extremely satisfied, 7 = extremely dissatisfied, 8 = not sure/no opinion).[33] This outcome will not be measured at 9 months follow-up.

Demographic and clinical characteristics

- Patient characteristics are only collected in the first questionnaire and include various demographic and clinical variables, including: age, sex, education level, duration of complaints, weight, height and co-morbidities.
- The risk of persistent disabling pain will be assessed with the Keele STarT MSK Tool (i.e. low, medium or high risk).[34] The Keele STarT MSK Tool is part of the Stratified Blended Approach and is additionally included in the data collection.
- As part of the data collection, patients' suitability for e-Exercise (blended care) will be measured by two self-developed questions as substitute for the Dutch Blended Physiotherapy Checklist.[35] It is not possible to use the Dutch Blended Physiotherapy Checklist as a measurement instrument, because it is a tool to guide physiotherapists in their clinical reasoning while setting up a personalized blended physiotherapy treatment,

thus not a patient reported outcome measure.[35] Therefore, this cannot be measured in the control arm and substitute questions will be used. The following questions will be assessed in the first questionnaire to measure smartphone or tablet access and usage: 'Do you own a smartphone or tablet? (yes/no)' and 'How many apps do you use regularly (weekly) on your smartphone or tablet? (none/1-3 per week/4-10 per week/more than 10 per week)'.

Other outcomes

- Compliance to the smartphone app with e-Exercise modules in the Stratified Blended Approach arm will be assessed by quantitative data on the usage. These data will automatically be stored on the backend of the app. Additionally, all patients will be asked in the first follow-up questionnaire whether they received and used an app or paper-based workbook as part of their physiotherapy treatment.

- The content and intensity of physiotherapy care will be described from trial case report forms, for both trial arms, filled out by the physiotherapists at the end of the treatment period or after 3 months. Information of the risk of persistent disabling pain, the suitability for blended care, the physiotherapists diagnosis of the presenting problem, the number of physiotherapy sessions, deviations from the study protocol, and content of the physiotherapy sessions will be collected. Number of physiotherapy sessions and content of the physiotherapy sessions will be presented in a table, separately for the 3 risk groups and per trial arm.

Cost-effectiveness

A cost-utility analysis (CUA) will be performed for QALYs and a cost-effectiveness analysis (CEA) for the combined region-specific pain and disability score, both of which will be performed from the societal (primary outcome) and the healthcare (secondary outcome) perspective. From the societal perspective all costs will be taken into account, irrespective of who pays or benefits, whereas solely those borne by the healthcare sector will be included if the healthcare perspective is applied.[36]

Identification, measurement and valuation of costs

Societal costs will be determined during 9 months of follow-up by gathering information on the patients' healthcare utilization, informal care, and (unpaid) productivity losses due to neck and/or shoulder complaints. This will be done by asking patients to complete three retrospective 3-monthly cost questionnaires. The costs of the Stratified Blended Approach will be estimated using a bottom-up micro costing approach.[37] Other kinds of healthcare utilization will include the use of primary care, secondary care, and medication, all of which will be assessed by the cost questionnaires and valued using Dutch standard costs.[36] If standard costs are unavailable, prices reported by professional organizations will be used. Unpaid productivity losses will be valued using a Dutch recommended shadow price.[36] Paid productivity losses comprise of both sickness absence and presenteeism (i.e. reduced productivity while at work). Sickness absence will be assessed

using a modified version of the iMTA Productivity Cost Questionnaire (iPCQ) and will be valued in accordance with the "Friction Cost Approach" (FCA), with a friction period of 12 weeks and gender-specific price weights.[36,38] The FCA assumes that production losses are confined to the "friction period" (i.e. time needed to replace a sick worker).[38] The participants' level of presenteeism will be measured using the "World Health Organization – Work Performance Questionnaire" as well as a modified version of the iPCQ, and will be valued using gender-specific price weights as well.[36,38–40]

Measurement and valuation of health-related quality of life

Health-related quality of life will also be measured with the EQ-5D-5L. This questionnaire measures five self-reported health domains (i.e. mobility, self-care, usual activities, pain/discomfort and anxiety/depression).[41] For the cost-utility analysis (CUA), EQ-5D-5L health states will be converted into utility values using the Dutch tariff.[42] Subsequently, Quality Adjusted Life Years (QALYs) will be estimated by multiplying the duration a patient spent in a certain health state by the utility value of that health state, using linear interpolation between measurement points.

Analysis methods

Descriptive statistics (e.g. means and proportions) will be used to describe the main characteristics of the clusters (physiotherapists) and trial population (patients). Characteristics of physiotherapists that will be reported are: number of physiotherapists, sex, age, specialization, years of experience working as a physiotherapist, employment status and physiotherapy practice size (where the physiotherapist is employed). The demographic and clinical variables of patients collected in the first questionnaire will be compared between the arms of the trial to investigate potential selection bias. Demographic and clinical baseline measurements of those who are and are not lost to follow-up will be compared to investigate selective attrition.

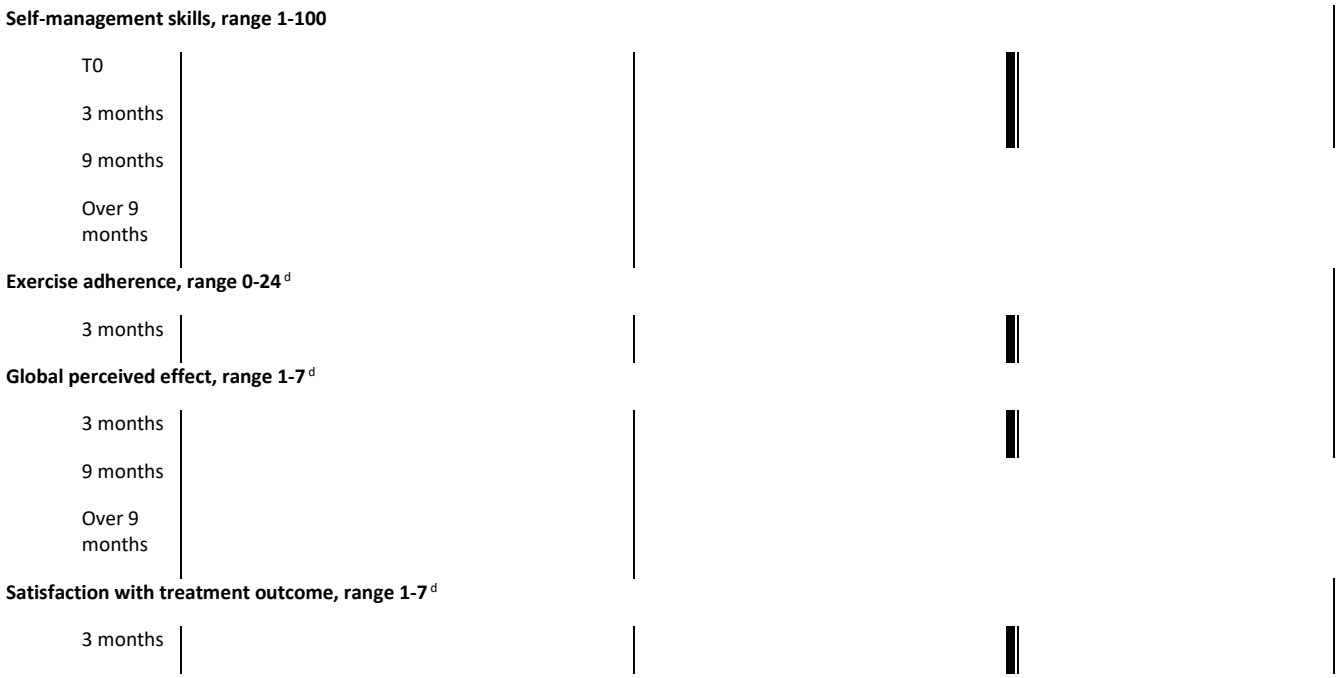
Effectiveness

To determine the overall effectiveness of the Stratified Blended Approach on the combined pain and disability score compared to usual physiotherapy in neck/shoulder patients over 9 months, differences in change scores per arm and time period will be estimated using linear mixed models (LMM) with random effects to control for correlation within patients and physiotherapists. Three levels are identified, consisting of repeated measurements (level 1), nested within patients (level 2), nested within physiotherapists (level 3). Analyses will be adjusted for the values at the first measurement and potential confounders: age, sex, type of complaints (neck or shoulder), pain intensity, duration of symptoms or other variables that are found to be imbalanced at the first measurement between arms.[43–48]

The statistical analysis of the primary outcome will also be used for the secondary outcomes. However, for dichotomous outcomes, a generalized mixed model (logit link) with the same multilevel structure will be used. Exploratory subgroup analyses will be carried out for hypotheses generating purposes. These analyses will be carried out to investigate potential differences in effectiveness within the three prognostic risk groups (low, medium or high risk), groups based on suitability for blended care (yes or no) and the neck and shoulder patient groups (self-reported predominantly neck or shoulder symptoms), compared to the usual physiotherapy arm.

Table 2. Unadjusted and Adjusted Primary and Secondary Outcome Measures: Improvements and Differences Within and Between arms.

Outcome measure	Stratified Blended Approach (n =)			Usual physiotherapy (n =)			Between-Arms[b]		
	n	Mean (95% CI)	Within-Arm Difference, Mean (95% CI) (*P*)[c]	n	Mean (95% CI)	Within-Arm Difference, Mean (95% CI) (*P*)[c]	Between-Arms Difference Mean (95% CI)	*(P)* [c]	Between-Arms Effect Size
Combined measure of pain and disability, range 0-100									
T0									
3 months									
6 months									
9 months									
Over 9 months									
Health Related Quality of Life, range 0-100									
T0									
3 months									
9 months									
Over 9 months									
Physical activity, MVPA[a] average minutes per day									
T0									
3 months									
9 months									
Over 9 months									
Pain Intensity, range 0-10									
T0									
3 months									
9 months									
Over 9 months									
Illness perceptions, range 0-10									
T0									
3 months									
9 months									
Over 9 months									



[a] MVPA = Moderate to Vigorous Physical Activity

[b] Difference between T0 and follow-up measurements in Stratified Blended Physiotherapy vs. Usual Physiotherapy

[c] Unadjusted for baseline confounders

[d] These measures could only be measured after treatment-period, not at T0

Cost-effectiveness

For the CUA and CEA, missing cost and effect data will be imputed using multivariate imputation by chained equations.[49] The results of the imputed datasets will be pooled using Rubin's rules.[49] LMM, with the same three-level structure as described above, will be performed to estimate cost and effect differences.[50] See the tables below. In order to account for the highly skewed nature of cost data, bias-corrected and accelerated bootstrapping with 5000 replications will be used to estimate 95% confidence intervals around the cost differences. Incremental Cost-Effectiveness Ratios (ICERs) will subsequently be calculated by dividing the differences in costs between trial arms by the difference in QALYs for the CUA and the differences in combined measure of pain and disability for the CEA. The uncertainty surrounding the ICERs will be graphically illustrated by plotting bootstrapped incremental cost-effect pairs on cost-effectiveness planes.[51] Moreover, cost-effectiveness acceptability curves (CEACs) will be constructed to provide a summary measure of the joint uncertainty of costs and effects. CEACs indicate the probability of the Stratified Blended Approach being cost-effective in comparison to usual physiotherapy care at different willingness-to-pay values.[52] To test the robustness of the trial results, several sensitivity analyses will be performed.

Tabel 3. Mean costs per participant in the Stratified Blended Approach arm and usual physiotherapy (PT) arm and mean differences between both arms during 9 months follow-up

	SBA (*N* =); mean costs in € (SEM)	Usual PT (*N* =); mean costs in € (SEM)	Unadjusted mean cost difference in € (95% CI)	Adjusted mean cost difference in € (95% CI)
Intervention [a]				
Primary healthcare				
Secondary healthcare				
Medication [a]				
Sport				
Informal care				
Absenteeism				
Presenteeism				
Unpaid productivity				
Healthcare costs[b]				
Total costs				

[a]Significant difference between Stratified Blended Approach and usual PT; [b]Healthcare costs = intervention costs + primary healthcare costs + secondary healthcare costs + medication costs adjusted for sex, age, BMI, level of education, type of complaints (neck or shoulder), physical functioning at baseline, pain at baseline and utility score at baseline

Table 4. Differences in pooled mean costs and effects

Analysis	N SBA	N Usual PT	Outcome	ΔC (95% CI) In €	ΔE (95% CI) In points	ICER Euro/point	Distribution CE-plane (%)			
							NE[a]	SE[b]	SW[c]	NW[d]
Main analysis 1:			QALYs (0–1)							
Total costs and imputed dataset			region-specific pain and disability score (0–100)							
Main analysis 2:			QALYs (0–1)							
Healthcare costs and imputed dataset			region-specific pain and disability score (0–100)							
Sensitivity analysis 1:			QALYs (0–1)							
Complete cases			region-specific pain and disability							

	score (0–100)
Sensitivity analysis 2:	QALYs (0–1)
Per-protocol and imputed dataset	region-specific pain and disability score (0–100)

CI confidence interval, C costs, CE-plane cost-effectiveness plane, E effects, ICER incremental cost-effectiveness Ratio costs are expressed in 2015 Euros; [a]The northeast quadrant of the CE plane, indicating that SBA is more effective and more costly than usual PT; [b]The southeast quadrant of the CE plane, indicating that e-Exercise is more effective and less costly than usual physiotherapy; [c]The northwest quadrant of the CE plane, indicating that SBA is less effective and more costly than usual PT; [d]The southwest quadrant of the CE plane, indicating that SBA is less effective and less costly than usual PT

References

1. van Tilburg ML, Kloek CJJ, Pisters MF, Staal JB, van Dongen JM, de Weerd M, et al. Stratified care integrated with eHealth versus usual primary care physiotherapy in patients with neck and/or shoulder complaints: protocol for a cluster randomized controlled trial. BMC Musculoskelet Disord. 2021;22(1).

2. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: Extension to cluster randomised trials. BMJ (Online). 2012;345(7881).

3. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the content of statistical analysis plans in clinical trials. JAMA - Journal of the American Medical Association. 2017;

4. Jorritsma W, Dijkstra PU, De Vries GE, Geertzen JHB, Reneman MF. Detecting relevant changes and responsiveness of Neck Pain and Disability Scale and Neck Disability Index. European Spine Journal. 2012;

5. Yao M, Xu B ping, Tian Z rui, Ye J, Zhang Y, Wang Y jun, et al. Cross-cultural adaptation of the Neck Pain and Disability Scale: a methodological systematic review. Spine Journal. Elsevier; 2019.

6. Jorritsma W, De Vries GE, Dijkstra PU, Geertzen JHB, Reneman MF. Neck Pain and Disability Scale and Neck Disability Index: Validity of Dutch language versions. European Spine Journal. 2012 Jan;21(1):93–100.

7. Blozik E, Himmel W, Kochen MM, Herrmann-Lingen C, Scherer M. Sensitivity to change of the Neck Pain and Disability Scale. European Spine Journal. 2011 Jun 8;20(6):882–9.

8. Jorritsma W, De Vries GE, Geertzen JHB, Dijkstra PU, Reneman MF. Neck Pain and Disability Scale and the Neck Disability Index: Reproducibility of the Dutch Language Versions. European Spine Journal. 2010 Oct 28;19(10):1695–701.

9. Thoomes-de Graaf M, Scholten-Peeters W, Duijn E, Karel Y, de Vet HCW, Koes B, et al. The Responsiveness and Interpretability of the Shoulder Pain and Disability Index. Journal of Orthopaedic & Sports Physical Therapy. 2017 Apr;47(4):278–86.

10. MacDermid JC, Solomon P, Prkachin K. The Shoulder Pain and Disability Index demonstrates factor, construct and longitudinal validity. BMC Musculoskelet Disord. 2006 Feb 10;7:12.

11. Thoomes-de Graaf M, Scholten-Peeters GGM, Schellingerhout JM, Bourne AM, Buchbinder R, Koehorst M, et al. Evaluation of measurement properties of self-administered PROMs aimed at patients with non-specific shoulder pain and “activity limitations”: a systematic review. Vol. 25, Quality of Life Research. 2016. p. 2141–60.

12. Roe Y, Soberg H, Bautz-Holter E, Ostensjo S. A systematic review of measures of shoulder pain and functioning using the International classification of functioning, disability and health (ICF). BMC Musculoskelet Disord. 2013;14(1):73.

13. Schmidt S, Ferrer M, González M, González N, Valderas JM, Alonso J, et al. Evaluation of shoulder-specific patient-reported outcome measures: A systematic and standardized comparison of available evidence. J Shoulder Elbow Surg. 2014 Mar 1;23(3):434–44.

14. Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. Pain. 1999;

15. Breivik EK, Björnsson GA, Skovlund E. A comparison of pain rating scales by sampling from clinical trial data. Clinical Journal of Pain. 2000;

16. Zee KI Van Der, Sanderman R. Het meten van de algemene gezondheidstoestand met de Rand-36: een handleiding. NCG reeks meetinstrumenten. 2012.

17. Zee KI, Sanderman R, Heyink JW, Haes H. Psychometric qualities of the rand 36-item health survey 1.0: A multidimensional measure of general health status. Int J Behav Med. 2005;

18. Aaronson NK, Muller M, Cohen PDA, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol. 1998;51(11):1055–68.

19. Bullinger M, Alonso J, Apolone G, Leplège A, Sullivan M, Wood-Dauphinee S, et al. Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol. 1998 Nov;51(11):913–23.

20. Ware JE, Gandek B, Kosinski M, Aaronson NK, Apolone G, Brazier J, et al. The Equivalence of SF-36 Summary Health Scores Estimated Using Standard and Country-Specific Algorithms in 10 Countries: Results from the IQOLA Project. J Clin Epidemiol. 1998 Nov 1;51(11):1167–70.

21. Leysen M, Nijs J, Meeus M, Paul van Wilgen C, Struyf F, Vermandel A, et al. Clinimetric properties of illness perception questionnaire revised (IPQ-R) and brief illness perception questionnaire (Brief IPQ) in patients with musculoskeletal disorders: A systematic review. Manual Therapy. 2015.

22. Hallegraeff JM, Van Der Schans CP, Krijnen WP, De Greef MHG. Measurement of acute nonspecific low back pain perception in primary care physical therapy: Reliability and validity of the brief illness perception questionnaire. BMC Musculoskelet Disord. 2013;

23. Broadbent E, Wilkes C, Koschwanez H, Weinman J, Norton S, Petrie KJ. A systematic review and meta-analysis of the Brief Illness Perception Questionnaire. Psychol Health. 2015;

24. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res. 2005;

25. Rademakers J, Nijman J, Van Der Hoek L, Heijmans M, Rijken M. Measuring patient activation in the Netherlands: Translation and validation of the American short form Patient Activation Measure (PAM13). BMC Public Health. 2012;

26. Robusto KM, Trost SG. Comparison of three generations of ActiGraph^{TM} activity monitors in children and adolescents. J Sports Sci. 2012;30(13):1429–35.

27. Aadland E, Ylvisåker E. Reliability of the actigraph GT3X+ accelerometer in adults under free-living conditions. PLoS One. 2015;10(8):e0134606.

28. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? International Journal of Behavioral Nutrition and Physical Activity. 2011;

29. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc. 1998;

30. Newman-Beinart NA, Norton S, Dowling D, Gavriloff D, Vari C, Weinman JA, et al. The development and initial psychometric evaluation of a measure assessing adherence to prescribed exercise: the Exercise Adherence Rating Scale (EARS). Physiotherapy (United Kingdom). 2017;

31. Kamper SJ, Ostelo RWJG, Knol DL, Maher CG, de Vet HCW, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol. 2010 Jul;63(7):760-766.e1.

32. Evans R, Bronfort G, Maiers M, Schulz C, Hartvigsen J. "I know it's changed": A mixed-methods study of the meaning of Global Perceived Effect in chronic neck pain patients. European Spine Journal. 2014 Apr;23(4):888–97.

33. Hudak PL, Wright JG. The characteristics of patient satisfaction measures. Spine. 2000.

34. Hill JC, Afolabi EK, Lewis M, Dunn KM, Roddy E, Van Der Windt DA, et al. Does a modified STarT Back Tool predict outcome with a broader group of musculoskeletal patients than back pain? A secondary analysis of cohort data. BMJ Open. 2016 Oct 14;6(10):e012445.

35. Kloek CJJ, Janssen J, Veenhof C. Development of a Checklist to Assist Physiotherapists in Determination of Patients' Suitability for a Blended Treatment. Telemed J E Health. 2020 Aug 1;26(8):1051–65.

36. Hakkaart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Swan Tan S. Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. Zorginstituut Nederland. 2016;

37. Frick KD. Microcosting Quantity Data Collection Methods. Med Care. 2009;

38. Koopmanschap MA, Rutten FFH, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ. 1995;

39. Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D, et al. The World Health Organization Health and Work Performance Questionnaire (HPQ). J Occup Environ Med. 2003;

40. Kessler RC, Ames M, Hymel PA, Loeppke R, McKenas DK, Richling DE, et al. Using the World Health Organization Health and Work Performance Questionnaire (HPQ) to Evaluate the Indirect Workplace Costs of Illness. J Occup Environ Med. 2004;

41. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy (New York). 1990;

42. Versteegh M, M. Vermeulen K, M. A. A. Evers S, de Wit GA, Prenger R, A. Stolk E. Dutch Tariff for the Five-Level Version of EQ-5D. Value in Health. 2016;19(4):343–52.

43. Artus M, Campbell P, Mallen CD, Dunn KM, Van Der Windt DAW. Generic prognostic factors for musculoskeletal pain in primary care: A systematic review. BMJ Open. 2017.

44. Verwoerd M, Wittink H, Maissan F, de Raaij E, Smeets RJEM. Prognostic factors for persistent pain after a first episode of nonspecific idiopathic, non-traumatic neck pain: A systematic review. Musculoskelet Sci Pract. 2019 Jul;42:13–37.

45. Kuijpers T, Van Der Windt DAWM, Van Der Heijden GJMG, Bouter LM. Systematic review of prognostic cohort studies on shoulder disorders. Pain. 2004;109(3):420–31.

46. Struyf F, Geraets J, Noten S, Meeus M, Nijs J. A multivariable prediction model for the chronification of non-traumatic shoulder pain: A systematic review. Vol. 19, Pain Physician. American Society of Interventional Pain Physicians; 2016. p. 1–10.

47. Bruls VEJ, Bastiaenen CHG, De Bie RA. Prognostic factors of complaints of arm, neck, and/or shoulder: A systematic review of prospective cohort studies. Vol. 156, Pain. Lippincott Williams and Wilkins; 2015. p. 765–88.

48. Paksaichol A, Janwantanakul P, Purepong N, Pensri P, Van Der Beek AJ. Office workers' risk factors for the development of non-specific neck pain: A systematic review of prospective cohort studies. Vol. 69, Occupational and Environmental Medicine. Occup Environ Med; 2012. p. 610–8.

49. Campion WM, Rubin DB. Multiple Imputation for Nonresponse in Surveys. Journal of Marketing Research. 2006;

50. El Alili M, van Dongen JM, Goldfeld KS, Heymans MW, van Tulder MW, Bosmans JE. Taking the Analysis of Trial-Based Economic Evaluations to the Next Level: The Importance of Accounting for Clustering. Pharmacoeconomics. 2020 Jul 30;1–15.

51. Black WC. The CE Plane. Medical Decision Making. 2007;

52. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves - Facts, fallacies and frequently asked questions. Health Econ. 2004;

Supplementary Table

Table 5. Characteristics per subgroup at T0

Variable	T0 value					
	Low Risk		Medium Risk		High Risk	
	App	Paper-based	App	Paper-based	App	Paper-Based
No. of respondents						
Age, years						
Sex, female (%)						
Education level						
Low						
Middle						
High						
Type of complaints (%)						
Predominantly Neck						
Predominantly Shoulder						
Duration of physical symptoms						
<3 weeks						
3 – 12 weeks						
>12 weeks						
BMI, kg/m2						
Risk of persistent disabling pain (using the STarT MSK Tool)* (%)						
Low						
Medium						
High						
Smartphone or tablet access, yes (%)						
Smartphone or tablet usage						
None						
1 – 3 apps per week						
4 – 10 apps per week						
>10 apps per week						
No. of comorbidities						
0						
1						
≥2						
Combined region-specific pain and disability score†						
Health-related quality of life‡						
MVPA (average minutes per day)§						
Pain intensity¶						
Illness perceptions#						
Self-management skills**						